CONTENTS

1 Two-Way Crossed Classification

Days to first germination of three varieties of carrot seed grown in two types of potting soil.

Soil	Variety		
Tpye	1	2	3
1	$y_{111} = 6$ $y_{112} = 10$ $y_{113} = 11$	$y_{121} = 13$ $y_{122} = 15$	$y_{131} = 14$ $y_{132} = 22$
2	$y_{211} = 12$ $y_{212} = 15$ $y_{213} = 19$ $y_{214} = 18$	$y_{221} = 31$	$y_{231} = 18$ $y_{232} = 9$ $y_{233} = 12$

This might be called "an unbalanced factorial experiment".

Sample sizes:

Soil type	Variety 1	2	3
1	$n_{11} = 3$	$n_{12} = 2$	$n_{13} = 2$
2	$n_{21} = 4$	$n_{22} = 1$	$n_{23} = 3$

In general we have

$i = 1, 2, \ldots, a$ levels for the first factor

$j = 1, 2, \ldots, b$ levels for the second factor

$n_{ij} > 0$ observations at the i-th level of the first factor and the j-th level of the second factor

We will restrict our attention to normal-theory Gauss-Markov models.

1.1 "Cell Means" Model

$$y_{ijk} = \mu_{ij} + \epsilon_{ijk}$$

where

$$\epsilon_{ijk} \sim NID(0, \sigma^2) \quad \begin{cases} i = 1, \ldots, a \\ j = 1, \ldots, b \\ k = 1, \ldots, n_{ij} \end{cases}$$

Clearly, $E(y_{ijk}) = \mu_{ij}$ is estimable if $n_{ij} > 0$.

Overall mean response:

Mean response at i-th level of factor 1, averaging across the levels of factor 2.

Mean response at j-th level of factor 2, averaging across the levels of factor 1

Contrasts of interest:
"main effects" for factor 1:

"main effects" for factor 2:

Conditional effects:

Interaction contrasts:

All of these contrasts are **estimable** when

$$n_{ij} > 0 \quad \text{for all } (i, j)$$

because

- $E(\bar{y}_{ij.}) = \mu_{ij}$
- Any linear function of estimable functions is estimable

1.2 An “Effects” Model

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ij}$$

where

$$\epsilon_{ijk} \sim NID(0, \sigma^2)$$
$$i = 1, 2, \ldots, a$$
$$j = 1, 2, \ldots, b$$
$$k = 1, 2, \ldots, n_{ij} > 0$$

$$\begin{bmatrix} y_{111} \\ y_{112} \\ y_{113} \\ y_{121} \\ y_{122} \\ y_{131} \\ y_{132} \\ y_{211} \\ y_{212} \\ y_{213} \\ y_{214} \\ y_{221} \\ y_{231} \\ y_{232} \\ y_{233} \end{bmatrix} = \begin{bmatrix} 1&1&0&1&0&0&1&0&0&0&0&0 \\ 1&1&0&1&0&0&1&0&0&0&0&0 \\ 1&1&0&1&0&0&1&0&0&0&0&0 \\ 1&1&0&0&1&0&0&1&0&0&0&0 \\ 1&1&0&0&1&0&0&1&0&0&0&0 \\ 1&1&0&0&0&1&0&0&1&0&0&0 \\ 1&1&0&0&0&1&0&0&1&0&0&0 \\ 1&0&1&1&0&0&0&0&0&1&0&0 \\ 1&0&1&1&0&0&0&0&0&1&0&0 \\ 1&0&1&1&0&0&0&0&0&1&0&0 \\ 1&0&1&1&0&0&0&0&0&1&0&0 \\ 1&0&1&0&1&0&0&0&0&0&1&0 \\ 1&0&1&0&0&1&0&0&0&0&0&1 \\ 1&0&1&0&0&1&0&0&0&0&0&1 \\ 1&0&1&0&0&1&0&0&0&0&0&1 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \gamma_{11} \\ \gamma_{12} \\ \gamma_{13} \\ \gamma_{21} \\ \gamma_{22} \\ \gamma_{23} \end{bmatrix} + \boldsymbol{\epsilon}$$

1.2.1 Baseline Restrictions

The resulting restricted model is

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijk}$$

where

$$\epsilon_{ijk} \sim NID(0, \sigma^2) \quad \begin{cases} i = 1, \ldots, a \\ j = 1, \ldots, b \\ k = 1, \ldots, n_{ij} \end{cases}$$

and

$$\begin{aligned} \alpha_a &= 0 \\ \beta_b &= 0 \\ \gamma_{ib} &= 0 \;\; \text{for all} \;\; i = 1, \ldots, a \\ \gamma_{aj} &= 0 \;\; \text{for all} \;\; j = 1, \ldots, b \end{aligned}$$

We will call these the “baseline” restrictions.

Soil Type	Variety 1	Variety 2	Variety 3	Soil Type Means
1	$\mu_{11} = \mu + \alpha_1 + \beta_1 + \gamma_{11}$	$\mu_{12} = \mu + \alpha_1 + \beta_2 + \gamma_{12}$	$\mu_{13} = \mu + \alpha_1$	$\mu + \alpha_1 + \frac{\beta_1+\beta_2}{3} + \frac{\gamma_{11}+\gamma_{12}}{3}$
2	$\mu_{21} = \mu + \beta_1$	$\mu_{22} = \mu + \beta_2$	$\mu_{23} = \mu$	$\mu + \frac{\beta_1+\beta_2}{3}$
Var. means	$\mu + \frac{\alpha_1}{2} + \beta_1 + \frac{\gamma_{11}}{2}$	$\mu + \frac{\alpha_1}{2} + \beta_2 + \frac{\gamma_{12}}{2}$	$\mu + \frac{\alpha_1}{2}$	

Interpretation:

$\mu =$

$\alpha_i =$

$\beta_j =$

$\gamma_{ij} =$

Matrix formulation:

Least squares estimation:

Comments:

Imposing a set of restrictions on the parameters in the "effects" model

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijk}$$

to obtain a model matrix with full column rank.

(i) Avoids the use of a generalized inverse in least squares estimation.

(ii) Is equivalent to choosing a generalized inverse for $\mathbf{b} = (\mathbf{X}^T\mathbf{X})^-\mathbf{X}^T\mathbf{y}$ in the unrestricted "effects" model.

(iii) Restrictions must involve "non-estimable" quantities for the unrestricted "effects" model.

(iv) Baseline restrictions using by SAS are
$\alpha_a = 0 \qquad \beta_b = 0$
$\gamma_{ib} = 0 \quad$ for all $i = 1, \dots, a$
$\gamma_{aj} = 0 \quad$ for all $j = 1, \dots, b$

(v) Baseline restrictions using by R are
$\alpha_1 = 0 \qquad \beta_1 = 0$
$\gamma_{i1} = 0 \quad$ for all $i = 1, \dots, a$
$\gamma_{1j} = 0 \quad$ for all $j = 1, \dots, b$

1.2.2 Σ-Restrictions

$$y_{ijk} = \omega + \gamma_i + \delta_j + \eta_{ij} + \epsilon_{ijk}$$

$$\mu_{ij} = E(y_{ijk})$$

where

$$\epsilon_{ijk} \sim NID(0, \sigma^2) \text{ and } \sum_{i=1}^{a} \gamma_i = 0 \qquad \sum_{j=1}^{b} \delta_j = 0$$

$$\sum_{i=1}^{a} \eta_{ij} = 0 \quad \text{for each } j = 1, \ldots, b$$

$$\sum_{j=1}^{b} \eta_{ij} = 0 \quad \text{for each } i = 1, \ldots, a$$

Interpretation:

$\omega =$

$\delta_j - \delta_k =$

Similarly,
$\gamma_1 - \gamma_2 =$

For a model that includes the Σ-restrictions:
$\eta_{ij} =$

Matrix formulation:

Least squares estimation:

If restrictions are placed on "non-estimable" functions of parameters in the unrestricted "effects" model, then

- The resulting models are reparameterizations of each other.
- $\hat{\mathbf{y}} = P_{\mathbf{X}}\mathbf{y}$

  $\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}} = (I - P_{\mathbf{X}})\mathbf{y}$

  $SSE = \mathbf{e}^T\mathbf{e} = \mathbf{y}^T(I - P_{\mathbf{X}})\mathbf{y}$

  $\hat{\mathbf{y}}^T\hat{\mathbf{y}} = \mathbf{y}^T P_{\mathbf{X}}\mathbf{y}$

  $SS_{\text{model}} = \mathbf{y}^T(P_{\mathbf{X}} - P_{\mathbf{1}})\mathbf{y}$

  are the same for any set of restrictions.
- The solution to the normal equations

  $$\mathbf{b} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$$

  and interpretations of the corresponding parameters will not be the same for all such sets of restrictions.

If you were to place restrictions on estimable functions of parameters in

$$y_{ijk} = \mu + \alpha_1 + \beta_j + \gamma_{ij} + \epsilon_{ijk}$$

then you would change

- rank($\mathbf{X}$)
- space spanned by the columns of $\mathbf{X}$
- $\hat{\mathbf{y}} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-}\mathbf{X}^T\mathbf{y}$ and OLS estimators of other estimable quantities.

Example 1.

In a study to examine the effect of 4 drugs on 3 experimentally induced diseases in dogs, each drug-disease combination was given to six randomly selected dogs. The measurement (y) to be analyzed was the increase in systolic blood pressure (mm Hg) due to treatment. Unfortunately, some dogs were unable to complete the experiment. The data are shown in the following table.

	Disease		
Drug	$j=1$	$j=2$	$j=3$
$i=1$	42, 44, 36,13, 19, 22	33, 26, 33,21	31, -3, 25,25, 24
$i=2$	28, 23, 24,42, 13	34, 33, 31,36	3, 26, 28,32, 3, 16
$i=3$	1, 29, 19	11, 9, 7,1, -6	21, 1, 9,3
$i=4$	24, 9, 22,-2, 15	27, 12, 12,-5, 16, 15	22, 7, 25,5, 12

Consider the model $y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijk}$ where $\epsilon_{ijk} \sim NID(0, \sigma^2)$ and y_{ijk} denotes the change in systolic blood pressure (mm Hg) for the $k-$th dog given the $j-$th disease and treated with the $i-$th drug.

(a) Note that the application of the lm() function in R imposes some restrictions to solve the normal equations. What are the restrictions?

(b) Using the solution to the normal equations provided by the application of the of lm() function in R, report estimates of the following quantities:

$$\mu,\ \alpha_1,\ \beta_3,\ \gamma_{23},\ \alpha_2 - \alpha_3,\ \mu + \alpha_2 + \beta_3 + \gamma_{23}$$

```
setwd("E:")
dogs <- read.table("dogs.dat", col.names=c("Drug","Disease","Y"))
dogs
dogs$Drug <- as.factor(dogs$Drug)
dogs$Disease <- as.factor(dogs$Disease)
options( contrasts=c("contr.treatment", "contr.ploy") )
lm.out1 <- lm( Y ~ Drug*Disease, data=dogs )
lm.out1$coef
> lm.out1$coef
   (Intercept)  Drug2  Drug3     Drug4       Disease2
    29.33     -3.33    -13.00    -15.73     -1.083
 Disease3 Drug2:Disease2 Drug3:Disease2 Drug4:Disease2 Drug2:Disease
    -8.93      8.583    -10.85           0.32           0.93
Drug3:Disease3 Drug4:Disease3
     1.10      9.53
```


(c) Give an interpretation of the following quantities.

$$\mu, \alpha_1, \beta_3, \gamma_{23}, \mu + \alpha_2 + \beta_3 + \gamma_{23}$$

(d) The are many ways to put linear restrictions on parameters in the original model to obtain a solution to the normal equations. Would the least squares estimates of any of the linear combinations of parameters in part (a) have the same value for all such solutions to the normal equations? Which ones? Explain.

1.3 Normal Theory Gauss-Markov Model

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijk}$$

1.3.1 Analysis of Variance

$$\begin{aligned}
\mathbf{y}^T\mathbf{y} &= \mathbf{y}^T P_\mu \mathbf{y} + \mathbf{y}^T (P_{\mu,\alpha} - P_\mu)\mathbf{y} \\
&\quad + \mathbf{y}^T (P_{\mu,\alpha,\beta} - P_{\mu,\alpha})\mathbf{y} \\
&\quad + \mathbf{y}^T (P_{\mathbf{X}} - P_{\mu,\alpha,\beta})\mathbf{y} \\
&\quad + \mathbf{y}^T (I - P_{\mathbf{X}})\mathbf{y} \\
&= R(\mu) + R(\boldsymbol{\alpha}|\mu) + R(\boldsymbol{\beta}|\mu, \alpha) \\
&\quad + R(\boldsymbol{\gamma}|\mu, \boldsymbol{\alpha}, \boldsymbol{\beta}) + SSE
\end{aligned}$$

By Cochran's Theorem, these quadratic forms (or sums of squares) have independent chi-square distributions with 1, $a-1$, $b-1$, $(a-1)(b-1)$, and $n_{\bullet\bullet} - ab$ degrees of freedom, respectively, (if $n_{ij} > 0$ for all (i, j))

$$
\begin{bmatrix} y_{111} \\ y_{112} \\ y_{113} \\ y_{121} \\ y_{122} \\ y_{131} \\ y_{132} \\ y_{211} \\ y_{212} \\ y_{213} \\ y_{214} \\ y_{221} \\ y_{231} \\ y_{232} \\ y_{233} \end{bmatrix}
=
\left[\begin{array}{c|cc|ccc|cccccc}
1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\
1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\
1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\
1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1
\end{array}\right]
\begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \gamma_{11} \\ \gamma_{12} \\ \gamma_{13} \\ \gamma_{21} \\ \gamma_{22} \\ \gamma_{23} \end{bmatrix}
+ \boldsymbol{\epsilon}
$$

$\uparrow$ call this $\mathbf{X}_\mu$ $\quad\nwarrow$ call this $\mathbf{X}_\alpha$ $\quad\nwarrow$ call this $\mathbf{X}_\beta$ $\quad\uparrow$ call this $\mathbf{X}_\gamma$

Define:

$$\mathbf{X}_\mu = \mathbf{X}_\mu \qquad P_\mu = \mathbf{X}_\mu(\mathbf{X}_\mu^T\mathbf{X}_\mu)^{-1}\mathbf{X}_\mu^T$$

$$\mathbf{X}_{\mu,\alpha} = [\mathbf{X}_\mu | \mathbf{X}_\alpha] \qquad P_{\mu,\alpha} = \mathbf{X}_{\mu,\alpha}(\mathbf{X}_{\mu,\alpha}^T\mathbf{X}_{\mu,\alpha})^{-}\mathbf{X}_{\mu,\alpha}^T$$

$$\mathbf{X}_{\mu,\alpha,\beta} = [\mathbf{X}_\mu | \mathbf{X}_\alpha | \mathbf{X}_\beta] \qquad P_{\mu,\alpha,\beta} = \mathbf{X}_{\mu,\alpha,\beta}\left(\mathbf{X}_{\mu,\alpha,\beta}^T\mathbf{X}_{\mu,\alpha,\beta}\right)^{-}\mathbf{X}_{\mu,\alpha,\beta}^T$$

$$\mathbf{X} = [\mathbf{X}_\mu | \mathbf{X}_\alpha | \mathbf{X}_\beta | \mathbf{X}_\gamma] \qquad P_{\mathbf{X}} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-}\mathbf{X}^T$$

The following three model matrices correspond to reparameterizations of the same model:

Mode1 1:

$$
\begin{bmatrix}
1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\
1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\
1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\
1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1
\end{bmatrix}
\begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \gamma_{11} \\ \gamma_{12} \\ \gamma_{13} \\ \gamma_{21} \\ \gamma_{22} \\ \gamma_{23} \end{bmatrix}
$$

Model 2:	Model 3:
$\left[\begin{array}{c\|ccccc} 1&1&1&0&1&0\\ 1&1&1&0&1&0\\ 1&1&1&0&1&0\\ 1&1&0&1&0&1\\ 1&1&0&1&0&1\\ 1&1&0&0&0&0\\ 1&1&0&0&0&0\\ 1&0&1&0&0&0\\ 1&0&1&0&0&0\\ 1&0&1&0&0&0\\ 1&0&1&0&0&0\\ 1&0&0&1&0&0\\ 1&0&0&0&0&0\\ 1&0&0&0&0&0\\ 1&0&0&0&0&0 \end{array}\right] \begin{bmatrix} \mu\\ \alpha_1\\ \beta_1\\ \beta_2\\ \gamma_{11}\\ \gamma_{12} \end{bmatrix}$	$\left[\begin{array}{c\|ccccc} 1&1&1&0&1&0\\ 1&1&1&0&1&0\\ 1&1&1&0&1&0\\ 1&1&0&1&0&1\\ 1&1&0&1&0&1\\ 1&1&-1&-1&-1&-1\\ 1&1&-1&-1&-1&-1\\ 1&-1&1&0&-1&0\\ 1&-1&0&1&-1&0\\ 1&-1&1&0&-1&0\\ 1&-1&1&0&-1&0\\ 1&-1&0&1&0&-1\\ 1&-1&-1&-1&1&1\\ 1&-1&-1&-1&1&1\\ 1&-1&-1&-1&1&1 \end{array}\right] \begin{bmatrix} \omega\\ \gamma_1\\ \delta_1\\ \delta_2\\ \eta_{11}\\ \eta_{12} \end{bmatrix}$

$R(\mu) = \mathbf{y}^T P_\mu \mathbf{y}$ is the same for all three models
$R(\mu, \boldsymbol{\alpha}) = \mathbf{y}^T P_{\mu,\alpha} \mathbf{y}$ is the same for all three models and so is $R(\alpha|\mu) = R(\mu, \boldsymbol{\alpha}) - R(\mu)$
$R(\mu, \boldsymbol{\alpha}, \boldsymbol{\beta}) = \mathbf{y}^T P_{\mu,\alpha,\beta} \mathbf{y}$ is the same for all three models and so is $R(\beta|\boldsymbol{\alpha}) = R(\mu, \boldsymbol{\alpha}, \boldsymbol{\beta}) - R(\mu, \boldsymbol{\alpha})$
$R(\mu, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}) = \mathbf{y}^T P_{\mathbf{X}} \mathbf{y}$ is the same for all three models and so is $R(\gamma|\mu, \boldsymbol{\alpha}, \boldsymbol{\beta}) = R(\mu, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}) - R(\mu, \boldsymbol{\alpha}, \boldsymbol{\beta})$

Consequently, the partition

$$\begin{aligned} \mathbf{y}^T\mathbf{y} &= \mathbf{y}^T P_\mu \mathbf{y} + \mathbf{y}^T (P_{\mu,\beta} - P_\mu)\mathbf{y} \\ &\quad + \mathbf{y}^T (P_{\mu,\alpha,\beta} - P_{\mu,\beta})\mathbf{y} \\ &\quad + \mathbf{y}^T (P_{\mathbf{X}} - P_{\mu,\alpha,\beta})\mathbf{y} \\ &\quad + \mathbf{y}^T (I - P_{\mathbf{X}})\mathbf{y} \\ &= R(\mu) + R(\boldsymbol{\beta}|\mu) + R(\boldsymbol{\alpha}|\mu, \boldsymbol{\beta}) \\ &\quad + R(\boldsymbol{\gamma}|\mu, \alpha, \beta) + SSE \end{aligned}$$

is the same for all three models.

By Cochran's Theorem, these quadratic forms (or sums of squares) have independent chi-square distributions with 1, $b-1$, $a-1$, $(a-1)(b-1)$, and $n_{\bullet\bullet} - ab$ degrees of freedom, respectively, when $n_{ij} > 0$ for all (i, j).

We have also shown earlier that

$$\begin{aligned} SSE &= \mathbf{y}^T(I - P_{\mathbf{X}})\mathbf{y} \\ &= \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n_{ij}}(y_{ijk} - \bar{y}_{ij\bullet})^2 \\ &\sim \chi^2_{n_{\bullet\bullet}-ab} \end{aligned}$$

Example 2.

Let $\mathbf{Y} \sim N(\mathbf{W}\boldsymbol{\gamma}, \sigma^2 I)$, where

- $\mathbf{W} = \begin{bmatrix}\mathbf{W_1} & \mathbf{W_2} & \mathbf{W_3} & \mathbf{W_4}\end{bmatrix}$,
- $\mathbf{W_1} = \mathbf{1_{20}}$,
- $\mathbf{W_2} = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \otimes \mathbf{1_{10}}$,
- $\mathbf{W_3} = \mathbf{1_2} \otimes \begin{bmatrix} 1 \\ -1 \end{bmatrix} \otimes \mathbf{1_5}$,
- $\mathbf{W_4} = \mathbf{1_4} \otimes \begin{bmatrix} -6 \\ -3 \\ 0 \\ 6 \\ 3 \end{bmatrix}$, and
- $\boldsymbol{\gamma} = \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \\ \gamma_4 \end{bmatrix}$

(a) Use Cochran's theorem to find the distributions of

- $\frac{1}{\sigma^2}SSE = \mathbf{e^Te} = \mathbf{Y^T(I - P_W)Y}$, where $\mathbf{P_W} = \mathbf{W(W^TW)^{-1}W^T}$
- $\frac{1}{\sigma^2}R(\gamma_1) = \mathbf{Y^TP_{W_1}Y}$ where $\mathbf{W_1} = \mathbf{1}$ is the first column of $\mathbf{W}$ and $\mathbf{P_{W_1}} = \mathbf{W_1(W_1^TW)^{-1}W_1^T}$.
- $\frac{1}{\sigma^2}R(\gamma_2|\gamma_1) = \mathbf{Y^T(P_{W_2} - P_{W_1})}Y$ where $\mathbf{W_2}$ contains the first two columns of $\mathbf{W}$ and $\mathbf{P_{W_2}} = \mathbf{W_2(W_2^TW_2)^{-1}W_2^T}$.
- $\frac{1}{\sigma^2}R(\gamma_3|\gamma_1\gamma_2) = \mathbf{Y^T(P_{W_3} - P_{W_2})Y}$. where $\mathbf{W_3}$ contains the first three columns of $\mathbf{W}$ and $\mathbf{P_{W_3}} = \mathbf{W_3(W_3^TW_3)^{-1}W_3^T}$.
- $\frac{1}{\sigma^2}R(\gamma_4|\gamma_1\gamma_2\gamma_3) = \mathbf{Y^T(P_W - P_{W_3})Y}$.

(b) Report a formula for the non-centrality parameter of the non-central F distribution of

$$F = \frac{R(\gamma_3|\gamma_1, \gamma_2)}{SSE/7}$$

Use it to the null and alternative hypotheses associated with this test statistic. You are given that:

$$\mathbf{W^T(P_{W_3} - P_{W_2})W}] = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 20 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

1.3.2 Type I Sum of Squares

What null hypotheses are tested by F-tests derived from such ANOVA tables

$R(\mu)=$

For the carrot seed germination study:

$$\begin{aligned}P_{\mathbf{1}}\mathbf{X}\boldsymbol{\beta} &= \frac{1}{n_{..}}\mathbf{1}\,\mathbf{1}^T\mathbf{X}\boldsymbol{\beta}\\ &= \frac{1}{n_{..}}\mathbf{1}[n_{..}, n_{1.}, n_{2.}, n_{.1}, n_{.2}, n_{.3},\\ &\quad n_{11}, n_{12}, n_{13}, n_{21}, n_{22}, n_{23}]\boldsymbol{\beta}\\ &= \frac{1}{n_{..}}\mathbf{1}\left(n_{..}\mu + \sum_{i=1}^{a} n_{i.}\alpha_i + \sum_{j=1}^{b} n_{.j}\beta_j\right.\\ &\quad \left. + \sum_{i=1}^{a}\sum_{j=1}^{b}\gamma_{ij}\right)\end{aligned}$$

The null hypothesis is

$$H_0 : n_{..}\,\mu + \sum_{i=1}^{a} n_{i.}\,\alpha_i + \sum_{j=1}^{b} n_{.j}\,\beta_j + \sum_i \sum_j n_{ij}\gamma_{ij} = 0$$

With respect to the cell means

$$E(y_{ijk}) = \mu_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij}$$

this null hypothesis is

$$H_0 : \sum_{i=1}^{a}\sum_{j=1}^{b} \frac{n_{ij}}{n_{..}}\mu_{ij} = 0$$

Consider

$R(\boldsymbol{\alpha}|\mu) =$

For the general effects model for the carrot seed germination study:

$$\begin{aligned} P_{\mu,\alpha}\mathbf{X} &= \mathbf{X}_{\mu,\alpha}(\mathbf{X}_{\mu,\alpha}^T\mathbf{X}_{\mu,\alpha})^{-}\mathbf{X}_{\mu,\alpha}^T\mathbf{X} \\ &= \mathbf{X}_{\mu,\alpha}\begin{bmatrix} n_{..} & n_{1.} & n_{2.} \\ n_{1.} & n_{1.} & 0 \\ n_{2.} & 0 & n_{2.} \end{bmatrix}^{-} \end{aligned}$$

$$\times \begin{bmatrix} n_{..} & n_{1.} & n_{2.} & n_{.1} & n_{.2} & n_{.3} & n_{11} & n_{12} & n_{13} & n_{21} & n_{22} & n_{23} \\ n_{1.} & n_{1.} & 0 & n_{11} & n_{12} & n_{13} & n_{11} & n_{12} & n_{13} & 0 & 0 & 0 \\ n_{2.} & 0 & n_{2.} & n_{21} & n_{22} & n_{23} & 0 & 0 & 0 & n_{21} & n_{22} & n_{23} \end{bmatrix}$$

$\downarrow$

$$= \mathbf{X}_{\mu,\alpha}\begin{bmatrix} 0 & 0 & 0 \\ 0 & \frac{1}{n_{1.}} & 0 \\ 0 & 0 & \frac{1}{n_{2.}} \end{bmatrix}\left[\qquad\qquad\right] =$$

$$\begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ \vdots & \vdots & \vdots \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ \vdots & \vdots & \vdots \\ 1 & 0 & 1 \end{bmatrix}\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & \frac{n_{11}}{n_{1.}} & \frac{n_{12}}{n_{1.}} & \frac{n_{13}}{n_{1.}} & \frac{n_{11}}{n_{..}} & \frac{n_{12}}{n_{..}} & \frac{n_{13}}{n_{..}} & 0 & 0 & 0 \\ 1 & 0 & 1 & \frac{n_{11}}{n_{1.}} & \frac{n_{22}}{n_{2.}} & \frac{n_{23}}{n_{2.}} & 0 & 0 & 0 & \frac{n_{21}}{n_{1.}} & \frac{n_{22}}{n_{2.}} & \frac{n_{23}}{n_{1.}} \end{bmatrix}$$

Then, the first seven rows of $(\mathbf{P}_{\mu,\alpha} - \mathbf{P}_{\mu})\mathbf{X}\boldsymbol{\beta}$ are

$$\left[\mu + \alpha_1 + \textstyle\sum_{j=1}^{b} \frac{n_{1j}}{n_{1.}} (\beta_j + \gamma_{1j})\right]$$

$$- \left[\mu + \sum_{i=1}^{a} \frac{n_{i.}}{n_{..}} \alpha_i + \sum_{j=1}^{b} \frac{n_{.j}}{n_{..}} \beta_j + \sum_{i} \sum_{j} \frac{n_{ij}}{n_{..}} \gamma_{ij}\right]$$

The last eight rows of $(P_{\mu,\alpha} - P_{\mu})\mathbf{X}\boldsymbol{\beta}$ are

$$\left[\mu + \alpha_2 + \textstyle\sum_{j=1}^{b} \frac{n_{2j}}{n_{2.}} (\beta_j + \gamma_{2j})\right]$$

$$- \left[\mu + \sum_{i=1}^{a} \frac{n_{i.}}{n_{..}} \alpha_i + \sum_{j=1}^{b} \frac{n_{.j}}{n_{..}} \beta_j + \sum_{i} \sum_{j} \frac{n_{ij}}{n_{..}} \gamma_{ij}\right]$$

The null hypothesis is

$$H_0 : \alpha_i + \sum_{j=1}^{b} \frac{n_{ij}}{n_{i.}} (\beta_j + \gamma_{ij})$$

are all equal $(i = 1, \dots, a)$

Consider $R(\boldsymbol{\beta}|\boldsymbol{\mu}, \boldsymbol{\alpha}) = \mathbf{y}^T (P_{\mu,\alpha,\beta} - P_{\mu,\alpha})\mathbf{y}$ and the corresponding F-statistic

$$F = \frac{R(\boldsymbol{\beta}|\mu, \boldsymbol{\alpha})/(b-1)}{MSE} \sim F_{(b-1, n_{..}-ab)}(\lambda)$$

Here,

$$\frac{1}{\sigma^2} R(\boldsymbol{\beta}|\boldsymbol{\mu}, \boldsymbol{\alpha}) \sim \chi^2_{\text{rank}(\mathbf{X}_{\mu,\alpha,\beta}) - \text{rank}(\mathbf{X}_{\mu,\alpha})}(\lambda)$$

$$[1+(a-1)+(b-1)] - [1+(a-1)]$$
$$= b - 1 \text{ degrees of freedom}$$

and

$$\lambda = \frac{1}{\sigma^2} \Big[(P_{\mu,\alpha,\beta} - P_{\mu,\alpha})\mathbf{X}\boldsymbol{\beta}\Big]^T \Big[(P_{\mu,\alpha,\beta} - P_{\mu,\alpha})\mathbf{X}\boldsymbol{\beta}\Big]$$

$$P_{\mu,\alpha,\beta}\mathbf{X} = \mathbf{X}_{\mu,\alpha,\beta}\left[\mathbf{X}^T_{\mu,\alpha,\beta}\mathbf{X}_{\mu,\alpha,\beta}\right]^{-}\mathbf{X}^T_{\mu,\alpha,\beta}\mathbf{X}$$

$$= \mathbf{X}_{\mu,\alpha,\beta}\left[\begin{array}{c|cc|ccc} n_{..} & n_{1.} & n_{2.} & n_{.1} & n_{.2} & n_{.3} \\ \hline n_{1.} & n_{1.} & 0 & n_{11} & n_{12} & n_{13} \\ n_{2.} & 0 & n_{2.} & n_{21} & n_{22} & n_{23} \\ \hline n_{.l} & n_{11} & n_{21} & n_{.1} & 0 & 0 \\ n_{.2} & n_{12} & n_{22} & 0 & n_{.2} & 0 \\ n_{.3} & n_{13} & n_{23} & 0 & 0 & n_{.3} \end{array}\right]^{-}\mathbf{X}^T_{\mu,\alpha,\beta}\mathbf{X}$$

call this $\begin{bmatrix} A & B \\ B^T & C \end{bmatrix}$

$$\begin{bmatrix} A & B \\ B^T & C \end{bmatrix}^{-1} = \begin{bmatrix} A^{-1} & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} A^{-1}B \\ I \end{bmatrix}[C - B^TA^{-1}B]^{-1}[-B^TA^{-1} \mid I]$$

$$= \begin{bmatrix} 0 & 0 \\ 0 & C^{-1} \end{bmatrix} + \begin{bmatrix} I \\ -C^{-1}B^T \end{bmatrix}[A - BC^{-1}B^T]^{-1}[I \mid -BC^{-1}]$$

$$= \begin{bmatrix} W & -WBC^{-1} \\ -C^{-1}B^TW & C^{-1} + C^{-1}B^TWBC^{-1} \end{bmatrix}$$

where $W = [A - BC^{-1}B^T]^{-1}$

The null hypothesis is

$$H_0: \sum_{i=1}^{a}\frac{n_{ij}}{n_{.j}}(\beta_j + \gamma_{ij}) - \sum_{i=1}^{a}\frac{n_{ij}}{n_{.j}}\left(\sum_{k=1}^{b}\frac{n_{ik}}{n_{i.}}(\beta_k + \gamma_{ik})\right) = 0$$

for all $j = 1, \ldots, b$

With respect to the cell means,

$$E(y_{ijk}) = \mu_{ij},$$

this null hypothesis is

$$H_0: \sum_{i=1}^{a}\frac{n_{ij}}{n_{.j}}\mu_{ij} - \sum_{i=1}^{a}\frac{n_{ij}}{n_{.j}}\left(\sum_{k=1}^{b}\frac{n_{ik}}{n_{i.}}\mu_{ik}\right) = 0$$

for all $j = 1, 2, \ldots, b$.

Consider

$$R(\boldsymbol{\gamma}|\mu, \boldsymbol{\alpha}, \boldsymbol{\beta}) = \mathbf{y}^T[P_{\mathbf{X}} - P_{\mu,\alpha,\beta}]\mathbf{y}$$

and the associated F-statistic

$$F = \frac{R(\boldsymbol{\gamma}|\mu, \boldsymbol{\alpha}, \boldsymbol{\beta})/[(a-1)(b-1)]}{MSE} \sim F_{(a-1)(b-1), n_{\cdot\cdot}-ab}(\lambda)$$

The null hypothesis is:

$$H_0: \ (\mu_{ij} - \mu_{i\ell} - \mu_{kj} + \mu_{k\ell}) = (\gamma_{ij} - \gamma_{i\ell} - \gamma_{kj} + \gamma_{k\ell}) = 0$$

for all (i, j) and (k, ℓ) .

ANOVA Summary:

Sums of Squares	Associated null hypothesis
$R(\mu)$	$H_0: \mu + \sum_{i=1}^{a} \frac{n_{i\cdot}}{n_{\cdot\cdot}}\alpha_i + \sum_{j=1}^{b} \frac{n_{\cdot j}}{n_{\cdot\cdot}}\beta_j + \sum_{i=1}^{a}\sum_{j=1}^{b} \frac{n_{ij}}{n_{\cdot\cdot}}\gamma_{ij} = 0$ $\left(\text{or } H_0: \sum_{i=1}^{a}\sum_{j=1}^{b} \frac{n_{ij}}{n_{\cdot\cdot}}\mu_{ij} = 0\right)$
$R(\boldsymbol{\alpha}\|\mu)$	$H_0: \alpha_i + \sum_{j=1}^{b} \frac{n_{ij}}{n_{i\cdot}}(\beta_j + \gamma_{ij})$ are equal $\left(\text{or } H_0: \sum_{j=1}^{b} \frac{n_{ij}}{n_{i\cdot}}\mu_{ij} \text{ are equal}\right)$
$R(\boldsymbol{\beta}\|\mu, \boldsymbol{\alpha})$	$H_0: \beta_j + \sum_{i=1}^{a} \frac{n_{ij}}{n_{\cdot j}}\gamma_{ij} = \sum_{i=1}^{a} \frac{n_{ij}}{n_{\cdot j}} \sum_{k=1}^{b} \frac{n_{ik}}{n_{k\cdot}}(\beta_k + \gamma_{ik})$ for all $j = 1, \ldots, b$ $\left(\text{or } H_0: \sum_{i=1}^{a} \frac{n_{ij}}{n_{\cdot j}}\mu_{ij} = \sum_{i=1}^{a} \frac{n_{ij}}{n_{\cdot j}} \sum_{k=1}^{b} \frac{n_{ik}}{n_{i\cdot}}\mu_{ik} \text{ for all } j = 1, \ldots, b\right)$
$R(\boldsymbol{\gamma}\|\mu, \boldsymbol{\alpha}, \boldsymbol{\beta})$	$H_0: \gamma_{ij} - \gamma_{kj} - \gamma_{i\ell} + \gamma_{k\ell} = 0$ for all (i, j) and (k, ℓ) (or $H_0: \mu_{ij} - \mu_{kj} - \mu_{i\ell} + \mu_{k\ell} = 0$ for all $(i, j)$and (k, ℓ)

Sums of Squares	Associated null hypothesis
$R(\mu)$	$H_0 : \mu + \sum_{i=1}^{a} \frac{n_{i.}}{n_{..}}\alpha_i + \sum_{j=1}^{b} \frac{n_{.j}}{n_{..}}\beta_j + \sum_{i=1}^{a}\sum_{j=1}^{b} \frac{n_{ij}}{n_{..}}\gamma_{ij} = 0$ $\left(\text{or } H_0 : \sum_{i=1}^{a}\sum_{j=1}^{b} \frac{n_{ij}}{n_{..}}\mu_{ij} = 0\right)$
$R(\boldsymbol{\beta}\|\mu)$	$H_0 : \beta_j + \sum_{i=1}^{a} \frac{n_{ij}}{n_{.j}}(\alpha_j + \gamma_{ij})$ are equal for all $j = 1, \ldots, b$ $\left(\text{or } H_0 : \sum_{i=1}^{a} \frac{n_{ij}}{n_{.j}}\mu_{ij} \text{ are equal for all } j = 1, \ldots, b\right)$
$R(\boldsymbol{\alpha}\|\mu, \boldsymbol{\beta})$	$H_0 : \sum_{j=1}^{b} \frac{n_{ij}}{n_{i.}}(\alpha_{ij} + \gamma_{ij}) = \sum_{j=1}^{b} \frac{n_{ij}}{n_{i.}} \sum_{k=1}^{a} \frac{n_{kj}}{n_{.j}}(\alpha_k + \gamma_{kj})$ for all $i = 1, \ldots, a$ $\left(\text{or } H_0 : \sum_{j=1}^{b} \frac{n_{ij}}{n_{i.}}\mu_{ij} = \sum_{j=1}^{b} \frac{n_{ij}}{n_{i.}}\left[\sum_{k=1}^{a} \frac{n_{kj}}{n_{.j}}\mu_{kj}\right] \text{ for all } i = 1, \ldots, a\right)$
$R(\boldsymbol{\gamma}\|\mu, \boldsymbol{\alpha}, \boldsymbol{\beta})$	$H_0 : \gamma_{ij} - \gamma_{kj} - \gamma_{i\ell} + \gamma_{k\ell} = 0$ for all (i, j) and (k, ℓ) $\left(\text{or } H_0 : \mu_{ij} - \mu_{kj} - \mu_{i\ell} + \mu_{k\ell} = 0 \text{ for all } (i, j) \text{ and } (k, \ell)\right)$

Soil Tpye	Variety 1	2	3
1	$y_{111} = 6$	$y_{121} = 13$	$y_{131} = 14$
	$y_{112} = 10$	$y_{122} = 15$	$y_{132} = 22$
	$y_{113} = 11$		
2	$y_{211} = 12$	$y_{221} = 31$	$y_{231} = 18$
	$y_{212} = 15$		$y_{232} = 9$
	$y_{213} = 19$		$y_{233} = 12$
	$y_{214} = 18$		

Type I sums of squares

```
R-Codes
#Type I Sum of Squares(A follows by B)
Y = c(6, 10, 11, 13,15,14,22,12,15,19,18,31,18,9,12)
xmu = rep(1,15)
xa1 = c(rep(1,7),rep(0,8))
xa2 = 1-xa1
xalpha = cbind(xa1, xa2)
xb1 = c(rep(1,3), rep(0,4), rep(1,4), rep(0,4))
xb2 = c(rep(0,3), rep(1,2), rep(0,6), 1, rep(0,3))
xb3 = c(rep(0,5), 1,1, rep(0,5),rep(1,3))
xbeta = cbind(xb1,xb2,xb3)
xab11 = xa1*xb1
xab12 = xa1*xb2
xab13 = xa1*xb3
xab21 = xa2*xb1
xab22 = xa2*xb2
```

```
xab23 = xa2*xb3
xgamma = cbind(xab11,xab12,xab13,xab21,xab22,xab23)
library(MASS)
Pmu = xmu%*%solve(t(xmu)%*%xmu)%*%t(xmu)
xma = cbind(xmu, xalpha)
Pma = xma%*%ginv(t(xma)%*%xma)%*%t(xma)
xmab = cbind(xmu, xalpha, xbeta)
Pmab = xmab%*%ginv(t(xmab)%*%xmab)%*%t(xmab)
X = cbind(xmu, xalpha, xbeta, xgamma)
PX = X%*%ginv(t(X)%*%X)%*%t(X)
In = diag(rep(1,15))
A1 = Pmu
A2 = Pma - Pmu
A3 = Pmab - Pma
A4 = PX - Pmab
A5 = In - PX
Rmu = t(Y)%*%A1%*%Y
Rma = t(Y)%*%A2%*%Y
Rma
Rmab = t(Y)%*%A3%*%Y
Rmabg =  t(Y)%*%A4%*%Y
SSE = t(Y)%*%A5%*%Y
MRmu = Rmu
MRma = Rma
MRmab = Rmab/2
MRmabg = Rmabg/2
MSE = SSE/9
Fmu = MRmu/MSE
Fa = MRma/MSE
Fb = MRmab/MSE
Fab = MRmabg/MSE
PVmu = 1-pf(Fmu,1,9)
PVa = 1-pf(Fa,1,9)
PVb = 1-pf(Fb,2,9)
PVab = 1-pf(Fab,1,9)
data.frame(Source = "Intercept", SS=Rmu, df = 1, MS = MRmu, F.Stat = Fmu,
```



```
p.value = PVmu)
data.frame(Source = "Soil",SS=Rma, df = 1, MS = MRma, F.Stat = Fa,
p.value = PVa)
data.frame(Source = "Variety",SS=Rmab, df = 2, MS = MRmab, F.Stat = Fb,
p.value = PVb)
data.frame(Source = "Interaction",SS=Rmabg, df = 2, F.Stat = Fab,
p.value = PVab)
data.frame(Source = "Error",SS=SSE, df = 9,MS = MSE)
#----------------------------------------
#Using lm() function
Y = c(6, 10, 11, 13,15,14,22,12,15,19,18,31,18,9,12)
FA = as.factor(c(1,1,1,1,1,1,1,2,2,2,2,2,2,2,2))
FB = as.factor(c(1,1,1,2,2,3,3,1,1,1,1,2,3,3,3))
mod.fit = lm(Y ~ FA*FB)
anova(mod.fit)
```

Source of variat.	d.f.	sums of squares	Mean square	F	p-value
"Soils"	$a-1=1$	$R(\boldsymbol{\alpha}\|\boldsymbol{\mu}) = 52.50$	52.5	3.94	.0785
"Var."	$b-1=2$	$R(\boldsymbol{\beta}\|\mu,\boldsymbol{\alpha}) = 124.73$	62.4	4.68	.0405
Inter-action	(a-1)(b-1) -2	$R(\boldsymbol{\gamma}\|\mu,\boldsymbol{\alpha},\boldsymbol{\beta}) = 222.76$	111.38	8.35	.0089
"Res."	$\Sigma\Sigma(n_{ij}-1)$ =9	$\mathbf{y}^T(I-P_{\mathbf{X}})\mathbf{y} = 120.00$	13.33		
Corr. total	$n_{..}-1=14$	$\mathbf{y}^T(I-P_1)\mathbf{y} = 520.00$			

```
#Type I Sum of Squares(B follows by A)
Y = c(6, 10, 11, 13,15,14,22,12,15,19,18,31,18,9,12)
xmu = rep(1,15)
xa1 = c(rep(1,7),rep(0,8))
xa2 = 1-xa1
xalpha = cbind(xa1, xa2)
xb1 = c(rep(1,3), rep(0,4), rep(1,4), rep(0,4))
xb2 = c(rep(0,3), rep(1,2), rep(0,6), 1, rep(0,3))
xb3 = c(rep(0,5), 1,1, rep(0,5),rep(1,3))
xbeta = cbind(xb1,xb2,xb3)
xab11 = xa1*xb1
xab12 = xa1*xb2
xab13 = xa1*xb3
xab21 = xa2*xb1
xab22 = xa2*xb2
xab23 = xa2*xb3
xgamma = cbind(xab11,xab12,xab13,xab21,xab22,xab23)
library(MASS)
Pmu = xmu%*%solve(t(xmu)%*%xmu)%*%t(xmu)
xmb = cbind(xmu, xbeta)
Pmb = xmb%*%ginv(t(xmb)%*%xmb)%*%t(xmb)
xmab = cbind(xmu, xalpha, xbeta)
Pmab = xmab%*%ginv(t(xmab)%*%xmab)%*%t(xmab)
X = cbind(xmu, xalpha, xbeta, xgamma)
PX = X%*%ginv(t(X)%*%X)%*%t(X)
In = diag(rep(1,15))
A1 = Pmu
A2 = Pmb - Pmu
A3 = Pmab - Pmb
A4 = PX - Pmab
A5 = In - PX
Rmu = t(Y)%*%A1%*%Y
Rma = t(Y)%*%A2%*%Y
Rma
```



```
Rmab = t(Y)%*%A3%*%Y
Rmabg =  t(Y)%*%A4%*%Y
SSE = t(Y)%*%A5%*%Y
MRmu = Rmu
MRma = Rma
MRmab = Rmab/2
MRmabg = Rmabg/2
MSE = SSE/9
Fmu = MRmu/MSE
Fa = MRma/MSE
Fb = MRmab/MSE
Fab = MRmabg/MSE
PVmu = 1-pf(Fmu,1,9)
PVa = 1-pf(Fa,1,9)
PVb = 1-pf(Fb,2,9)
PVab = 1-pf(Fab,1,9)
data.frame(Source = "Intercept", SS=Rmu, df = 1, MS = MRmu, F.Stat = Fmu,
p.value = PVmu)
data.frame(Source = "Soil",SS=Rma, df = 1, MS = MRma, F.Stat = Fa,
p.value = PVa)
data.frame(Source = "Variety",SS=Rmab, df = 2, MS = MRmab, F.Stat = Fb,
p.value = PVb)
data.frame(Source = "Interaction",SS=Rmabg, df = 2, F.Stat = Fab,
p.value = PVab)
data.frame(Source = "Error",SS=SSE, df = 9,MS = MSE)
```

Source of variat.	d.f.	sums of squares	Mean square	F	p-value
"Var."	$b-1=2$	$R(\boldsymbol{\beta}\mid\mu)=93.33$	46.67	3.50	.0751
"Soils"	$a-1=1$	$R(\alpha\mid\boldsymbol{\mu},\boldsymbol{\beta})=83.90$	83.90	6.29	.0334
Inter-action	(a-1)(b-1) =2	$R(\boldsymbol{\gamma}\mid\mu,\boldsymbol{\alpha},\boldsymbol{\beta})=222.76$	111.38	8.35	.0089
"Res."	$\Sigma\Sigma(n_{ij}-1)$ =9	$\mathbf{y}^T(I-P_{\mathbf{X}})\mathbf{y}=120.00$	13.33		
Corr. total	$n_{\cdot\cdot}-1=14$	$\mathbf{y}^T(I-P_{\mathbf{1}})\mathbf{y}=520.00$			

1.3.3 Method of Unweighted Means - Type III Sum of Squares

(Type III sums of squares in when $n_{ij} > 0$ for all (i,j)).

Use the cell means reparameterization of the model:

$$\begin{aligned} y_{ijk} &= \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijk} \\ &= \mu_{ij} + \epsilon_{ijk} \end{aligned}$$

$$\underbrace{\begin{bmatrix} y_{111} \\ y_{112} \\ y_{113} \\ y_{121} \\ y_{122} \\ y_{131} \\ y_{132} \\ y_{211} \\ y_{212} \\ y_{213} \\ y_{214} \\ y_{221} \\ y_{231} \\ y_{232} \\ y_{233} \end{bmatrix}}_{\mathbf{y}} = \underbrace{\begin{bmatrix} 1&0&0&0&0&0 \\ 1&0&0&0&0&0 \\ 1&0&0&0&0&0 \\ 0&1&0&0&0&0 \\ 0&1&0&0&0&0 \\ 0&0&1&0&0&0 \\ 0&0&1&0&0&0 \\ 0&0&0&1&0&0 \\ 0&0&0&1&0&0 \\ 0&0&0&1&0&0 \\ 0&0&0&1&0&0 \\ 0&0&0&0&1&0 \\ 0&0&0&0&0&1 \\ 0&0&0&0&0&0 \\ 0&0&0&0&0&0 \end{bmatrix}}_{D} \underbrace{\begin{bmatrix} \mu_{11} \\ \mu_{12} \\ \mu_{13} \\ \mu_{21} \\ \mu_{22} \\ \mu_{23} \end{bmatrix}}_{\boldsymbol{\mu}} + \boldsymbol{\epsilon}$$

The model is

$$\mathbf{y} = D\boldsymbol{\mu} + \boldsymbol{\epsilon}$$

The least squares estimator (b.l.u.e.) for $\boldsymbol{\mu}$ is

Test the hypotheses are:

The OLS estimator (b.l.u.e.) for $\frac{1}{b}\sum_{j=1}^{b}\mu_{ij}$ and it's variance are

Express the null hypothesis in matrix form:

$$H_0 : C_1\boldsymbol{\mu} = \mathbf{0}$$

Then the OLS estimator (BLUE) of $C_1\boldsymbol{\mu}$, and it's variance are:

Compute SS_{H_0} and show that

$$\frac{1}{\sigma^2}SS_{H_0} \sim \chi^2_{(a-1)}(\lambda)$$

Compute:
$SSE = \mathbf{y}^T(I - P_D)\mathbf{y}$ where $P_D = D(D^TD)^{-1}D^T$.
Show that

$$\frac{1}{\sigma^2}SSE \sim \chi^2_{\Sigma\Sigma(n_{ij}-1)}$$

Show that

$$SSE = \mathbf{y}^T\underbrace{(I - P_D)}_{\nwarrow \text{ call this } A_1}\mathbf{y}$$

is distributed independently of

$$SS_{H_0} = \mathbf{y}^T\underbrace{D(D^TD)^{-1}C_1^T[C_1(D^TD)^{-1}C_1^T]^{-1}C_1(D^TD)^{-1}}_{\nwarrow \text{ call this } A_2}D^T\mathbf{y}$$

Then $F =$

Test

$$H_0 : \frac{1}{a}\sum_{i=1}^{a} \mu_{i1} = \frac{1}{a}\sum_{i=1}^{a} \mu_{i2} = \cdots = \frac{1}{a}\sum_{i=1}^{a} \mu_{ib}$$

vs.

$$H_A : \frac{1}{a}\sum_{i=1}^{a} \mu_{ij} \neq \frac{1}{a}\sum_{i=1}^{a} \mu_{ik} \quad \text{for some } j \neq k$$

Write the null hypothesis in matrix form as

$$H_0 : C_2\boldsymbol{\mu} = \mathbf{0}$$

where $C_2 =$

then $C_2\boldsymbol{\mu} =$

Compute $SS_{H_{0,2}}$ and reject H_0 if $F =$

Test for Interaction:

Test

$H_0 : \mu_{ij} - \mu_{i\ell} - \mu_{kj} + \mu_{k\ell} = 0$ for all (i, j) and (k, ℓ) vs.

$H_A : \mu_{ij} - \mu_{i\ell} - \mu_{kj} + \mu_{k\ell} \neq 0$ for all (i, k) and $(j \neq \ell)$.

Write the null hypothesis in matrix form as

$$H_0 : C_3\,\boldsymbol{\mu} = \mathbf{0}$$

and perform the test.


```
#Type III Sum of Squares
Y = c(6,10,11,13,15,14,22,12,15,19,18,31,18,9,12)
Y = c(6,10,11,13,15,14,22,12,15,19,18,31,18,9,12)
d1 = c(rep(1,3), rep(0,12))
d2 = c(0,0,0,1,1,rep(0,10))
d3 = c(rep(0,5),1,1,rep(0,8))
d4 = c(rep(0,7),rep(1,4),rep(0,4))
d5 = c(rep(0,11), 1, rep(0,3))
d6 = c(rep(0,12), 1, 1,1)
D = cbind(d1,d2,d3,d4,d5,d6)
a = 2
b = 3
beta  = solve(t(D)%*%D)%*%t(D)%*%Y
Yhat = D%*%beta
SSE = crossprod(Y-Yhat)
df2 = NROW(Y) - a*b
am1 = a-1
bm1 = b-1
Iam1 = diag(rep(1,am1))
Ibm1 = diag(rep(1,bm1))
Onea = c(rep(1,a))
Oneam1 = c(rep(1,am1))
Oneb = c(rep(1,b))
Onebm1 = c(rep(1,bm1))
C1 = kronecker(cbind(Iam1, -Oneam1),t(Oneb))
C1b = C1%*%beta
SSH0a = t(C1b)%*%
solve(C1%*%solve(crossprod(D))%*%t(C1))%*%C1b
df1 = b-1
```

```
F = (SSH0a/df1)/(SSE/df2)
p = 1-pf(F, df1,df2)
C1
data.frame(SS=SSH0a, df = df1, F.Stat = F, p.value = p)

C2 = kronecker(t(Onea), cbind(Ibm1, -Onebm1))
C2b = C2%*%beta
SSH0b = t(C2b)%*%
solve(C2%*%solve(crossprod(D))%*%t(C2))%*%C2b
df1 = b-1
F = (SSH0b/df1)/(SSE/df2)
p = 1-pf(F, df1,df2)
C2
data.frame(SS=SSH0b, df = df1, F.Stat = F, p.value = p)

C3 = kronecker(cbind(Iam1, -Oneam1), cbind(Ibm1, -Onebm1))
C3b = C3%*%beta
SSH0ab = t(C3b)%*%
solve(C3%*%solve(crossprod(D))%*%t(C3))%*%C3b
df1 = (a-1)*(b-1)
F = (SSH0ab/df1)/(SSE/df2)
p = 1-pf(F, df1,df2)
C3
data.frame(SS=SSH0ab, df = df1, F.Stat = F, p.value = p)
```


Source of variation	Sum of d.f.	Mean Squares	Square	F	p-value
Soils	a-1=1	$SS_{H_0} =$ 123.77	123.77	9.28	.0139
Var.	b-1=2	$SS_{H_{0,2}} =$ 192.13	96.06	7.20	.0135
Inter.	(a-1)(b-1)=2	$SS_{H_{0,3}} =$ 222.76	111.38	8.35	.0089

Note that

$$\begin{aligned}\mathbf{y}^T P_{\mathbf{1}}\mathbf{y} &+ \mathbf{y}^T D(D^T D)^{-1}[C_1(D^T D)^{-1}C_1^T]^{-1} \\ &\quad C_1(D^T D)^{-1}D^T\mathbf{y} \\ &+ \mathbf{y}^T D(D^T D)^{-1}C_2^T[C_2(D^T D)^{-1}C_2^T]^{-1} \\ &\quad C_2(D^T D)^{-1}D^T\mathbf{y} \\ &+ \mathbf{y}^T D(D^T D)^{-1}C_3^T[C_3(D^T D)^{-1}C_3^T]^{-1} \\ &\quad C_3(D^T D)^{-1}D^T\mathbf{y} \\ &+ \mathbf{y}^T(I - P_D)\mathbf{y}\end{aligned}$$

do not necessarily sum to $\mathbf{y}^T\mathbf{y}$, nor do the middle three terms (SS_{H_0}, $SS_{H_0,2}$, $SS_{H_0,3}$) necessarily sum to

$$SS_{\text{model,corrected}} = \mathbf{y}^T(P_D - P_{\mathbf{1}})\mathbf{y} ,$$

nor are (SS_{H_0}, $SS_{H_0,2}$, $SS_{H_0,3}$) necessarily independent of each other.

Example 3.

A chemical production process consists of a first reaction with an alcohol and a second reaction with a base. A 3×2 factorial experiment with three alcohols and two bases was conducted. The data had unequal replications among the six treatment combinations of the two factors, Base and Alcohol. The collected data are percent yield. The data are given below.

	Alcohol		
Base	1	2	3
1	90.1 91.5	89.2 88.2 90.5	90.5 87.9 89.2 91.2
2	88.5 88.6 92.0	96.0	93.4 90.8 92.8

Consider the model $y_{ijk} = \mu_{ij} + \epsilon_{ijk}$, where $\epsilon_{ijk} \sim NID(0, \sigma^2)$, $i = 1, 2$, and $j = 1, 2, 3$ and $k = 1, \ldots, n_{ij}$. This model can be expressed in matrix form as $\mathbf{Y} = \mathbf{D}\boldsymbol{\beta} + \boldsymbol{\epsilon}$. Examine type III sums of squares for these data.

(a) Specify the $\mathbf{C}$ matrix needed to write the null hypothesis associated with the F-test for Alcohol effects in the form $H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{0}$.

(b) Present a formula for SS_{H_0} , corresponding to the null hypothesis in part (a), and state it's distribution when the null hypothesis is true.

(c) Compute SS_{H_0}.

1.4 Balanced Factorial Experiments

$$n_{ij} = n \quad \text{for} \quad \begin{array}{l} i = 1, \ldots, a \\ j = 1, \ldots, b \end{array}$$

Example 4. Sugar Cane yields

	Nitrogem Level		
	150 lb/acre	210 lb/acre	270 lb/acre
Variety 1	$y_{111} = 70.5$	$y_{121} = 67.3$	$y_{131} = 79.9$
	$y_{112} = 67.5$	$y_{122} = 75.9$	$y_{132} = 72.8$
	$y_{113} = 63.9$	$y_{123} = 72.2$	$y_{133} = 64.8$
	$y_{114} = 64.2$	$y_{124} = 60.5$	$y_{134} = 86.3$
Variety 2	$y_{211} = 58.6$	$y_{221} = 64.3$	$y_{231} = 64.4$
	$y_{212} = 65.2$	$y_{222} = 48.3$	$y_{232} = 67.3$
	$y_{213} = 70.2$	$y_{223} = 74.0$	$y_{233} = 78.0$
	$y_{214} = 51.8$	$y_{224} = 63.6$	$y_{234} = 72.0$
Variety 3	$y_{311} = 65.8$	$y_{321} = 64.1$	$y_{331} = 56.3$
	$y_{312} = 68.3$	$y_{322} = 64.8$	$y_{332} = 54.7$
	$y_{313} = 72.7$	$y_{323} = 70.9$	$y_{331} = 66.2$
	$y_{314} = 67.6$	$y_{324} = 58.3$	$y_{334} = 54.4$

For a balanced experiment ($n_{ij} = n$), Type I, Type II, and Type III sums of squares are the same:

$R(\boldsymbol{\alpha}|\mu) =$

$R(\boldsymbol{\beta}|\mu) =$

$R(\boldsymbol{\gamma}|\mu, \boldsymbol{\alpha}, \boldsymbol{\beta}) =$

ANOVA

Sum of Squares	Associated null hypothesis
$R(\mu) = \mathbf{y}^T P_{\mathbf{1}}\mathbf{y}$	$H_0 : \mu + \frac{1}{a}\sum_{i=1}^{a}\alpha_i + \frac{1}{b}\sum_{j=1}^{b}\beta_j$
$= a\,b\,n\,\bar{y}_{...}^2$	$+ \frac{1}{a\,b}\sum_{i=1}^{a}\sum_{j=1}^{b}\gamma_{ij} = 0$
	$\left(H_0 : \frac{1}{a\,b}\sum_{i=1}^{a}\sum_{j=1}^{b}\mu_{ij} = 0\right)$
$R(\boldsymbol{\alpha}\|\mu) = R(\boldsymbol{\alpha}\|\mu, \boldsymbol{\beta})$	$H_0 : \alpha_i + \frac{1}{b}\sum_{j=1}^{b}(\beta_j + \gamma_{ij})$
$= n\,b\sum_{i=1}^{a}(\bar{y}_{i..} - \bar{y}_{...})^2$	are equal
	$\left(H_0 : \frac{1}{b}\sum_{j=1}^{b}\mu_{ij} \text{ are equal}\right)$
$R(\boldsymbol{\beta}\|\mu) = R(\boldsymbol{\beta}\|\mu, \boldsymbol{\alpha})$	$H_0 : \beta_j + \frac{1}{a}\sum_{i=1}^{a}(\alpha_i + \gamma_{ij})$
$= n\,a\sum_{j=1}^{b}(\bar{y}_{.j.} - \bar{y}_{...})^2$	are equal
	$\left(H_0 : \frac{1}{a}\sum_{i=1}^{a}\mu_{ij} \text{ are equal}\right)$

$$R(\boldsymbol{\gamma}|\mu,\boldsymbol{\alpha},\boldsymbol{\beta}) = n\sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{y}_{ij.}-\bar{y}_{i..}-\bar{y}_{.j.}+\bar{y}_{...})^2$$

$$H_0 : \gamma_{ij} - \gamma_{kj} - \gamma_{i\ell} + \gamma_{k\ell} = 0$$
$$\text{for all } (i, j) \text{ and } (k, \ell)$$

$$\Big(H_0 : \mu_{ij} - \mu_{kj} - \mu_{i\ell} + \mu_{k\ell} = 0$$
$$\text{for all } (i, j) \text{ and } (k, \ell) \Big)$$